DOCUMENT RESUME

ED 036 748 AC 006 554

AUTHOR Steele, Sara M.; Moss, Gwenna M.
TITLE The Criteria Problem in Program Evaluation.
PUB DATE Feb 70
NOTE 37p.; Paper presented at the Adult Education Research Conference, Minneapolis, Minnesota, February 27-28, 1970

EDRS PRICE EDRS Price MF-$0.25 HC-$1.95
DESCRIPTORS *Adult Education Programs, *Decision Making, *Evaluation Criteria, Evaluation Needs, *Evaluation Techniques, Evaluative Thinking, Program Design, *Program Evaluation, Research

ABSTRACT

The nature and role of criteria in evaluation of adult education programs are presented in this study. The first half of the paper discusses several premises: The goal of evaluation is to use program judgments in the program decision making process. Criteria are essential to judgment and determining the type of criteria that is relevant is an important part of the evaluation process. Criteria may function at various levels of specificity. Effectiveness of evaluation rests mainly with the quality of the criteria used; and efficiency in evaluation is determined by the evaluator's competency in developing criteria. The second half of the paper asserts that criteria are not adequately understood and used. Reasons suggested for this include: (1) too narrow a concept of evaluation; (2) too great an emphasis on information and program description; (3) the wish to avoid decision making; (4) the wish to escape challenge by avoiding subjectivity; and (5) poorly defined guides to data interpretation. Suggestions for improving the use of criteria are given. (author/pt)

ED036748

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

THE CRITERIA PROBLEM IN PROGRAM EVALUATION*

One of the major reasons why program evaluation often flounders is that too little attention is being given to a very crucial element--that of criteria. Recent adult education text books seldom mention the term in their chapters on program evaluation.[1] Although the term does appear in the literature on curriculum evaluation in other fields it is considered self-explanatory.[2] Little discussion is given to criteria as a phenomenon or to their role in evaluation. Adult educators are much more accustomed to applying criteria to products involved in their teaching--essays, paintings, cakes and cows--than they are to applying them to their own programs.

It is the point of this paper than an understanding and use of criteria is essential to program evaluation. The first part of the paper explores some of the ideas that we believe are basic to an understanding of criteria. The second part discusses some of the reasons why criteria are the "Cinderella's" of evaluation literature and outlines ways of improving their use.

Evaluation has been defined in many ways. We find the following definition most relevant:

> Evaluation is the systematic process of judging the worth, desirability, effectiveness, or adequacy of something according to definite criteria and purposes. The judgment is based upon a careful comparison of observation data with criteria standards. Precise definitions of what is to be appraised, clearly-stated purposes, specific standards for the criteria traits, accurate observations and measurements and logical conclusions are hallmarks of valid evaluation.[3]

Program evaluation in adult education involves two very complex concepts--the complexity of that which is adult education programs and the complexity which is evaluation. The adult educator must accept

---

*Sara M. Steele, Associate Professor, and Gwenna M. Moss, Instructor, Department of Agricultural and Extension Education, University of Wisconsin. Paper presented at the 1970 Adult Education Research Conference, Minneapolis, Minnesota, February, 1970.

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY Sara M. Steele
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF

this complexity and come to grips with it. Over-simplification results in confusion and frustration and utlimately in valueless evaluation.

The Nature and Role of Criteria in Program Evaluation

The term criteria as used in this paper means:

> measures against which something can be judged. They may be rules, standards, norms, objects, or conditions of behavior which are considered to be "good" or "ideal". They provide a description or image of what a valuable program is like, thereby assigning value to given phenomenon related to the program.[4]

Notice that the term criteria is used rather than criterion. This is deliberate. There are many types of criteria and many specific criteria which can be useful in program evaluation. The adult educator needs to recognize and deal with the multiplicity of appropriate criteria rather than seeking a few absolute statements of criterion.

The understanding of criteria has been more completely developed in evaluation situations which focus on people (teacher and personnel evaluation) and on things such as plants and systems (school accreditation) than it has in terms of evaluating programs, and yet many of the same basic ideas apply.

The following seven premises give an overview of some of the things that need to be understood if criteria are to be useful tools in evaluation:

1. The goal of evaluation is the use of program judgments in the program decision making process. The evaluation and the criteria used must be trusted by all relevant decision makers.

2. Evaluation is not complete unless judgment occurs. Criteria are essential to judgment.

3. Determining the type of criteria that is relevant is an important part of the process of doing evaluation. Selection of type should be governed by the kinds of decisions that need to be made about the program. Understanding alternative criteria categories helps to focus evaluation.

4. Criteria have both a conceptual and a performance component and may function at various levels of specificity.

5. The rational approach to criteria formation is the most relevant in the typical program evaluation situation.

6. The effectiveness of evaluation rests primarily with the quality of the criteria used.

7. Efficiency in evaluation is determined, to a great extent, by the evaluator's competency in developing criteria that are crucial and critical to his purpose in evaluation.

We will briefly sketch some of the basic ideas involved in each of the premises. It is not the purpose of this paper to cover any one premise exhaustively.

Premise 1. The goal of evaluation is the use of program judgments in the program decision-making process. The evaluation and the criteria used must be trusted by all relevant decision-makers.

The current emphasis on evaluation as a major input into program decision making is a very healthy trend.[5] Evaluation has been called the "science of providing information for decision-making."[6]

In any programming situation there may be many decision makers.[7] Evaluation must be concerned with the input needed not only by the programmer and the administrator of his agency, but also with the needs of local, state, or national power structures when relevant. The needs of present or potential program participants for information secured from program evaluation must be considered if the programmer accords them actual decision making power.

In institutions which involve multiple decision makers, it is essential that the criteria used in program evaluation be trusted by all of the different decision makers.[8] When there is disagreement about what criteria are to be used, a good deal can be learned by an attempt to identify and deal with differences.[9] "Part of the responsibility of evaluation is to make known which standards are held by whom."[10]

Premise 2. Evaluation is not complete unless judgment occurs. Criteria are essential to judgment.

It is becoming generally recognized that description alone does not constitute evaluation. Judgment is also essential if evaluation is to be complete.[11]

We suggest that the foundation of a theory of evaluation lies in the very simple proposition that there are three major elements in evaluation: criteria, evidence and judgment.[12]

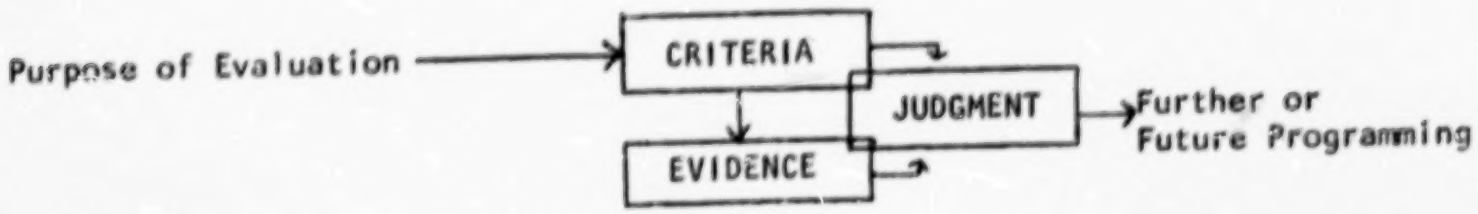

Figure 1.

The relationship of criteria to the other two elements illustrates the controlling role which they play in evaluation:

1. Criteria provide the framework within which evidence is collected. They tell us what to look for in the program or in the learner's behavior. They thus prescribe the nature of the essential evidence that needs to be obtained.
2. Criteria provide the base of comparison, or standards, which enable judgment to take place. Without these standards, judgment cannot occur, and if judgment does not occur, evaluation does not take place.

The role of criteria may be somewhat different in absolute judgement (one program compared to set of standards) and in relative judgment (two programs compared with criteria serving as the means of comparison),[13] but in all instances where judgments are being made there must be something (criteria) which serves as the basis for judgment. In casual

evaluation, the criteria may not even be at a conscious level. However, one of the key characteristics of systematic evaluation is the fact that the criteria are carefully developed and well thought through.

Premise 3. Determining the type of criteria that is relevant is an important part of the process of doing evaluation. Selection of type should be governed by the kinds of decisions that need to be made about the program. Understanding alternative criteria categories helps to focus evaluation.

A program can be judged against many types of criteria. The type is determined by the kind of decisions that must be made about the program. One of the first steps in the evaluation process is to determine the type or types of criteria that will be used.

The evaluator needs to have an understanding of the spectrum of criteria categories which can be applied to programs. There is a need for frameworks which provide a panoramic view. We are experimenting with a typology of criteria choices in and attempt to explore the dimensions of such a spectrum. A brief resume of it is presented here (even though it is in a very embryotic state) to illustrate the complexity of the choices that are involved in selecting the right kind of criteria for the particular decision.[14]

We suggest that the main selection of criteria categories should be done in terms of the characteristic of the program that is to be examined. We suggest that there are several different (although sometimes overlapping) characteristics that can be examined singly or in combination. Effectiveness, which includes the extent to which the program attains its objectives, is one such characteristic. (See Figure 2 for some others).

There are some criteria applying to a given characteristic which may remain constant regardless of other variables. Other criteria may be added to deal with programs in given situations. We suggest that such

Figure 2

Examples of Program Characteristics to be Considered in Evaluation

CHARACTERISTIC	EXAMPLES OF RELEVANT QUESTIONS	JUDGMENT TO BE MADE: HOW --	ILLUSTRATION OF INTENT
EFFORT	How much time was spent on this program? How many sessions were taught? etc.	Sufficient?	How many times did the bird flap his wings regardless of how far he flies?[a]
CONTACT	How many people were reached? communities? What proportion of the potential? What type of people? How much income was earned?	Sufficient?	How much air did the wings displace as they were flapping?
QUALITY	What was the quality of the content? learning experiences? media? environment? teacher's performance?	High?	How was the bird's coordination? Was the muscle formation adequate? The wing shaped right?
SUITABILITY	Did it meet the needs and expectations of the participants? of the community? Was it at the appropriate level? Was it within the unit's mission?	Great?	Was this the right bird to make this particular flight?
EFFECTIVENESS	What were the results? Did it accomplish its objectives? What were its effects?	Adequate?	How far has the bird actually flown? How far in terms of the total distance?
EFFICIENCY	Were the accomplishments sufficient for the amount of resources expended? Was this the best use of resources?	Sufficient?	Could the bird have arrived at his destination more efficiently by some other means? Did he take advantage of air currents? fly too high or not high enough?
IMPORTANCE	How valuable was it to: participants? to a greater society? Was its importance sufficient for the resources used? How much of total need was met?	Great?	How important was the bird's trip this time as compared with other possible trips?

a. Examples given by Suchman in regard to his categories of effort, performance, adequacy of performance and efficiency. pp. 61-65.

criteria may vary with program variables and with other variables. (See Figure 3.)

Among the program variables we include categories such as:

1. the intent of the program (e.g., congnitive or affective change.)
2. the stage of the programming cycle (design, in-process, wrap-up).
3. the approach used (group, one-to-one, media, etc.) and the method or technique (TV, work group, programmed instruction, demonstration, etc.).
4. whether the program process or a product of that process is to be examined.

Among the other variables are included:

1. characteristics of the clientele (criteria for a program for school drop outs may differ from those for lawyers).
2. situational factors (criteria for programs in situations where there are severe economic or cultural limitations on the participants' actions may have to differ from those where the social and cultural factors are in harmony with the direction of the program).
3. agency resources (criteria for situations where the agency has an abundance of resources may differ from those where the agency has meager resources.)

The latter three categories are those which are evaluated in pre-program evaluation.[15] They are also factors affecting criteria for most program characteristics. If the program decisions that are pending require a judgment of the suitability of a design for a program which is attempting to change attitudes of professional people who are firmly opposed to the attitude object and if that program is to be carried on by media, then the

FIGURE 3 - TYPOLOGY OF CHOICES IN PROGRAM EVALUATION
Major Choices - Panoramic View

Pre-Program Evaluation (program determination) →

OTHER CATEGORIES		
Clientele Characteristics	Situational Factors	Agency Resources

Program Evaluation

PROGRAM CATEGORIES			
Process Stage	Characteristic	Nature of Program	
	Type / Constant Criteria	Intent of the Program (See Fig. 3A)	Approaches and Techniques
Design	Effort		
	Contact		
	Quality		
	Suitability		
	Effectiveness		
	Importance		
	Efficiency		
	COMBINATION		
In-Process	Effort		
	Contact		
	Suitability		
	Quality		
	Effectiveness		
	Importance		
	Efficiency		
	COMBINATION		
Wrap-Up	Effort		
	Contact		
	Suitability		
	Quality		
(See Fig. 3B)	Effectiveness		
	Importance		
	Efficiency		
	COMBINATION		

Figure 3A Nature of the Program

CHARACTERISTIC	INTENT OF THE PROGRAM[a]				
	EXPOSURE	COGNITIVE CHANGE (Taxonomy I)	AFFECTIVE CHANGE (Taxonomy II)	CHANGE IN SKILL OR PRACTICE	DEVELOPMENTAL CHANGES	ENVIRONMENTAL CHANGES
DESIGN STAGE						
Effort						
Contact						
Quality						
Suitability						
Effectiveness						
Importance						
Efficiency						
Combination of the above						
IMPLEMENTATION STAGE						
Effort						
Contact						
etc.						
WRAP-UP STAGE						
Effort						
Contact						
etc.						

a. Each column might be further sub-divided if there are major differences in terms of the changes involved in the particular type of program. For example:

Cognitive Change					
			Use in:		
Recall	Comprehand	Apply	Analysis	Synthesis	Evaluation

or

Affective Change				
Receiving	Responding	Attending	Valuing	Value Set

or

Skill or Practice				
Perceiving	Set	Guided Response	Habitual Response	Automatic Response

Figure 3B

Characteristic of the Program -- Effectiveness

Effects Upon:[a]	Planned?		Number[b]		Duration[b]		
	In Objectives	Side Effects	Unitary	Multiple	Immediate	Intermediate[c]	Ultimate
PARTICIPANT	[d]						
His Family							
Peers							
Community and/or Society							
Programmer							
Other Programmers							
Agency							
Tax Payers							

a. Adapted from Scriven

b. Adapted from Suchman, p. 67

c. Adapted from Ryans

d. These are the only cells in this whole model that are considered evaluation in the very narrow interpretation of Tyler's concept of evaluation.

evaluator has clues as to the categories which he will be using in selecting or developing the criteria to be applied to the program. Different kinds of criteria would be applied in judging the effectiveness of a discussion group approach with inner city residents using knowledge in analyzing local problems.

In formative and summative evaluation,[16] the evaluator must select the category or categories of criteria which are relevant to the particular program decision. Usually only a few categories are used. In interpretive (seeking reasons)[17] evaluation and in evaluative research, the evaluator is concerned with the contingencies and congruencies between and among various sets of criteria and the program measurements that result. (e.g., Does a high quality design contribute to the effectiveness of the completed program?)

Because some have the feeling that one evaluates either according to objectives or according to criteria, we must point out that recent authors have shown that the attainment of objectives cannot be determined unless there are criteria for judging the behavior specified in those objectives.[18] In addition, we would suggest that there may be several other types of criteria which relate to program effectiveness. Several writers are now calling for the consideration of positive and negative side-effects, effects other than those stated in the objectives.[19] Scriven suggests that the effects upon relevant others (family, peers, the agency, tax payers, etc.) should be considered.[20] The evaluator is faced with choices within the category of effectiveness. Which effects? Upon whom?

The kind of criteria determine the nature of the evidence collected and the kind of judgment that can be made. The selection of the appropriate categories is an essential part of the evaluation process.

Premise 4. Criteria have both a conceptual and a performance component and may function at various levels of specificity.

Once the criteria category is determined the task of selecting or developing the set of criteria for the given evaluation problem begins.

Like objectives, criteria can be stated and function at various levels of specificity. The little rhyme cited by Herzog serves as a reminder:

> Big criteria have little criteria upon their backs to bite 'em. The small ones have still smaller, and so on ad infinitum.[21]

There is a growing ability to arrange objectives in hierarchial order in terms of specificity.[22] We need to develop the same ability in regard to criteria. Suchman's point of the relationships of assumptions of validity at different levels may well apply to criteria.[23]

In each category there may be several criterion elements--aspects which can be broken down into a set of criteria statements which relate to one overall criterion. When the criterion elements have been identified, each needs to be examined in terms of its structure--what its dimensions may be, what the various parts are like, and how they combine into meaningful patterns. The evaluator is concerned with developing a parsimonious description and then, if it is not possible to examine all of the parts, he is concerned with developing a sample of the criterion's dimensions which will adequately represent the whole criterion element.[24]

To be operational criteria must have two components. Each statement must include both a conceptual and a performance element. The conceptual criterion is a verbal statement of importance of certain outcomes or qualities. The performance criterion is any observable event which is

judged to be relevant to the conceptual criterion. The conceptual component is needed in order to understand the performance and "because the conceptual criterion usually implies something more than the actual criteria performance it may prove to be a valuable source either of future criterion measures or of improvements of existing ones."[25] It may be useful to think of this distinction in terms of a similar distinction from the research process. The conceptual criterion can be thought of as similar to the formal or theoretical definition of a variable, while the criterion performance is analagous to the working or operational definition of that variable.

The performance component of the criterion directs the means that will be used to measure the degree to which the criterion has been met. In some instances the criteria statements are converted directly into instruments by the addition of scales. In other instances there are intermediary steps between criteria statements and the development of instruments.

Level of criteria is an important concept when working from the general criteria category down to the specific statements that will be applied to the program. Level is also important in another sense. The evaluator is usually hoping to arrive at one overall summary judgment of the program (of ir not with one, with a very limited number). However, for criteria to be most useful they must be specific. There may be as many as ten or 15 small single judgments involved in the judgment of one criterion element. The evaluator needs to develop summary criteria at a higher level for one criterion element or for a combination of elements.

We call the level of the individual statements the micro-level and the level at which the many statements are summarized into a few overall judgments the macro-level. The macro-level criterion outlines the weights that will be given to the various sub-parts of the criteria and how they will be combined to come to one summary judgment of the program.[26] One might decide, for example, that a program will be considered fairly effective if 70 percent or more of the participants attain three out of the five objectives. Or one might decide that the program would be considered effective if 80 percent attain the first objectives and at least 40 percent attain each of the other three.

The conceptual and the performance elements of the criterion are equally important at the macro- and at the micro-levels. The statements at both levels must be specific enough to be operationalized with a good degree of reliability.

Premise 5. The rational approach to criteria development is the most relevant in the typical program evaluation situation.

Since criteria play such a central role in the evaluation process, the approach used to develop and/or select criteria therefore becomes of crucial importance.

A fundamental point to be recognized and accepted is that the selection and development of criteria statements is ultimately a matter involving value judgments.[27] They describe what the relevant decision makers value in a program. Once this point is accepted, then criteria development should proceed in a manner which takes this fact into account and operates to maximize the rationality of those judgments. Consequently this premise reaffirms Astin's statement that "criterion development is fundamentally a rational, non-empirical procedure,"[28] and advocates the use of the rational approach, as described by Ryans, as the primary approach to use.

In discussing the rational approach it will first be useful to distinguish it from the two other approaches: the arm-chair approach and the empirical approach.[29]

The arm-chair approach is described as "unanalyzed retrospective impressions, based upon non-systematic observation and often characterized by free-associations, and therefore likely to result in incomplete and contaminated descriptions of criteria."[30] The criteria which result from the arm-chair approach are highly subject to both intentional and unintentional selection bias. However, many of the thousands of criteria for teacher effectiveness, school housing, textbooks, salary schedules and vocational success are developed in this way.

The empirical approach "is a pragmatic one and consists essentially of 'trying out' hypothesized descriptions of the criterion, . . . and accepting, modifying, or rejecting the criterion framework in light of experience (e.g., intercorrelation data and evidences growing out of the application of sampling statistics)."[31] We would suggest that while the empirical approach does have a role to play in criteria development, this role is appropriately played only <u>after</u> criteria have been developed using the rational approach.

The rational approach "is centered in systematic observation and the logical analysis of the criterion behavior and its products, leading to an inclusion and exclusive designation of the components of the standard or base to be employed in making comparisons." It is systematic and comprehensive. "It aims to result in a description based on the relevance of possible criterion components, judged from the standpoint of belongingness and representative sampling."[32]

An important part of the "systematic analysis" inherent in the rational approach consists of a careful consideration of the sources from which criteria may be developed. The origin of lists of criteria appearing in the literature is usually unexplained. Because they are labelled as criteria one often assumes that they have some well-founded base. This isn't always the case. In general, criteria may be drawn from such sources as:

a. Accepted Principles. There are, however, few such generally accepted principles in the field of adult education.

b. Theory. Guidelines which are posited in place of principles.

c. Other studies. Including research, evaluation, and descriptive studies. Evaluation studies of similar programs can be a fruitful source. Similarly descriptions of other programs and their results (i.e. non-evaluative studies) can be analyzed for their potential contribution to criteria.

d. Practical Experience. Personal experience and that of others can be a rich source of criteria, particularly if those experiences are analyzed specifically for the purpose of criteria deriviation.

e. Philosophy. Since one's philosophy of education, or social action, is based on values of what should be achieved, criteria development will inevitably reflect this philosophy. It may be useful to explicitly take this into account by using one's philosophy as a screen through which criteria are filtered, much as Tyler[33] proposes the use of one's philosophy of education as a screen through which objectives pass.

In summary, when criteria are selected or developed for use in evaluation, the selection rests with the arbitrary judgment of the

evaluator. Recognizing this, the evaluator will want to be as rational as possible in his approach and select or develop the criteria with care. In addition he will need to understand the possible biasing effects of his past experiences and philosophy. Systematic analysis of the situation surrounding the criteria, and a consideration of a variety of sources of criteria will be useful. The rigor involved in logically justifying the inclusion of some criteria and the exclusion of others may also improve the process of criteria development.

Premise 6. The effectiveness of evaluation rests primarily with the quality of the criteria used.

In light of the controlling role which criteria play in evaluation, this premise directs attention to the quality of criteria as a necessary condition for effective evaluation.

Discussions of the quality of criteria generally revolve around considerations of their validity. At the outset it seems necessary to distinguish between the validity of criteria and the validity of the measurements based on those criteria. In regard to the latter the literature in research methodology is the appropriate source. We will not address ourselves to this aspect except to note that too much emphasis on measurement can be dangerous, since it focuses attention away from the central concern--the criteria. Improving the validity and reliability of measurements is of little use if the criteria upon which they are based are invalid. What is important here is a consideration of some of the problems involved in "validating" criteria as distinct from "validating" the measurements of those criteria, even though it is recognized that the two are to some extent interdependent.[34]

If Premise 5 is accepted, then the quality (or validity) of criteria is largely determined by the quality of the judgment used in

selecting criteria. Basically this implies that the validity of criteria is primarily a matter of attending to logical, face, or content validity. Empirical validation plays a subsidiary role to logical validation which involves the logical justification of the relevance of the criteria.[35]

Consequently improving the quality of criteria will be most dependent on improving the quality of the judgments employed as the basis for logically 'validating' criteria. Several suggestions made in our discussion of Premise 5 have relevance for this discussion (e.g. consideration of many sources, awareness of personal biases, etc.). In addition, it is worth drawing attention to several suggestions in the literature. These include:

a. Ryan's discussion of the use of a panel of authorities in making judgments on criteria.[36] Suchman refers to this as "consensual validity."[37]

b. Brogden and Taylor's classification of the factors biasing criteria, in particular their discussion of criterion deficiency and contamination.[38]

c. The use of a modification of Scriven's "consistency analysis" to obtain judgments as to the cohesiveness or relevance of criteria performance to conceptual criteria, ultimately extending to the measurements.[39]

The quality of the total evaluation will depend on the quality of the criteria. If their quality is high, the evaluation may attain high quality; if their quality is low the evaluation cannot attain high quality. In other words, the selection of quality criteria is a necessary although not sufficient condition for effective evaluation.

Premise 7. Efficiency in evaluation is determined, to a great extent, by the evaluator's competency in developing criteria that are CRUCIAL and CRITICAL to his purpose in evaluation.

The previous premise emphasized the quality of criteria in terms of their logical relevance and their inclusiveness and exclusiveness in relation to the conceptual criterion. This premise also emphasizes quality, but with reference to the need for criteria which are crucial and critical.

In a field like adult education where there has been little research on criteria, but where there is a good deal of theory, philosophy, and description, the evaluator could, if he proceeded in accord with the ideas in previous premises, come up with a large number of criteria which are relevant to his evaluation purposes. This pool of criteria could be drawn from a variety of sources, represent a range of types of criteria, and be characterized by a fairly complete coverage of his conceptual criteria.

However, for evaluation to be efficient in use of resources and judgments made, it will usually be necessary for the evaluator to limit the number of criteria which he develops or includes. It is suggested that the delimitation be based on a consideration of which criteria are the most crucial and critical for the purposes of the evaluation. Criteria which are crucial are those which absolutely must be included if the decisions called for in the evaluation are to be made. The omission of crucial criteria could well result in invalid judgments simply because a necessary aspect of the conceptual criterion was missing. Consideration of critical criteria revolves around sorting out those which make a real difference in the judgment process. In a sense this can be thought of as somewhat analagous to the ability of a measure to discriminate. The

need is for criteria which discriminate sufficiently among the judgment alternatives to allow a clear decision to be made.

The discussions of methods of weighting criteria which appear in the literature are essentially predicated on the assumption that some criteria are more important than others.[40] This assumption of differential importance is equally applicable whether weighting is used to derive composite scores or not.

As to the methods by which one arrives at critical and crucial criteria, again this would seem to call for rational decisions. While empirical measures can make a greater contribution here, over-reliance on them may mask an explicit consideration of the relative importance of criteria. A formal procedure (e.g. averaged estimates of judged importance by a panel of judges) has merit in making those decisions explicit.[41]

It may be worth noting that some critical criteria may relate to unintended negative side-effects.[42] Regardless of the presence of positive results on some criteria, a concommitant result in a negative direction might be considered critical enough to alter the judgment made.

In summarizing, it may be useful to think of the search for crucial and critical criteria as being analagous to the pruning of a tree. A tree in its natural state may grow in all directions but produce a smaller harvest of uneven quality. Pruning which leaves the healthiest branches growing in the right direction may result in a more bountiful harvest of higher quality.

THE PROBLEM: SOURCES AND SOLUTIONS

From the preceding premises it is apparent that there may be many problems involved in the development and use of criteria. The references cited have discussed several from a research point of view. The title of this paper, however, deliberately focuses on one problem. That problem is the fact that criteria are not presently being adequately used in program evaluation. In this section we will explore a few of the contributors to the lack of understanding and ability to use criteria. Then we will suggest some means of attacking the criteria problem.

Sources of the Problem

The following seven contributing factors do not exhaust the list of reasons why there is a lack of understanding and use of criteria. However, we feel they are some of the major contributors.

1. Too narrow a concept of evaluation

For several years now the basic concept of program evaluation held by many in the field of adult or extension education has been that of a wedding of the Tyler concept of evaluating the attainment of objectives[43] with traditional research methodology. This approach places most emphasis on evidence and gives little attention to criteria and judgment. The Tyler approach was a marked improvement over the measurement approach that was prevalent at the time of its inception. However, times are changing. Programming decisions are much more complex. We need a concept of evaluation and a theoretical framework that is broad enough to be meaningful in the typical decison making situation of today. We need new approaches to program evaluation. The objectives plus evidence approach is important but is only one of several important parts of a concept of program evaluation.

In the field of adult education, major limitations in the commonly held objectives plus evidence approach include: (1) the exclusion of other important judgments (for example those dealing with the suitability, importance, and efficiency of the program) which are extremely relevant to many program decisions; (2) the fact that the approach has little meaning if the programmers involved do not understand behavior sufficiently well to use objectives meaningfully in their programming. (Writing an objective does not mean that it is either realistic or that the teacher is able to actually teach for the attainment of that objective); (3) the degree of focus on pre-set objectives for a entire group of people. This is a carry over from elementary and secondary education which may not always be appropriate in adult education settings. It is particularly difficult to reconcile with a belief in the autonomy of the adult and the role of adult education in helping him achieve self-actualization.

There is a growing feeling in other areas of education that the objectives plus evidence concept of evaluation is too narrow for the job that needs to be done. Gruba comments[44] on some of the specific lacks related to this and the measurement approach. After considerable experience in evaluating new federally funded programs he feels that we are failing in program evaluation because we: lack an adequate definition of evaluation and adequate evaluation theory; lack knowledge about decisons processes; lack ceiteria; lack approaches differentiated by levels; lack mechanisms for organizang, processing and reporting evaluative information; and lack trained personnel. Most of these lacks are directly related to the nature, role, and ability to use criteria. He feels that the equating of the methodology of educational evaluation with the methodology of research has

led to some disastrous consequences.

2. Administrators are more interested in descriptions of results than in systematically derived judgments to be used internally.

Administrators in some adult education agencies (Cooperative Extension in p rticular) do encourage program evaluation. However, they often do so with a specific end in mind. They want evidence that can be shaped into reports for law makers and for supporting publics; they want information to tell the story of what the agency is doing. They are not as interested in information and criteria for making program decisions, even though such information and criteria could also be valuable in external reporting.

One sometimes suspects that there is a reverse principle operating in the administration of some adult education agencies. The more important the program decision the fewer the criteria and the fewer the relevant facts used in the decision.

For criteria to become an important part of program evaluation at the operational level, administration must provide leadership in their development and use. Furthermore, they must provide the kind of leadership that makes criteria a meaningful tool rather than an obstacle to be overcome or a challenge to be circumvented. Most of us are familiar with the games people play when criteria are not used in a meaningful way. It is often possible to comply with the words and completely avoid the intent. Like any other tool, criteria will only be as valuable as the way in which they are used.

3. The lack of a framework for guiding criteria development and interpretation.

In borrowing from research methodology, what many evaluators have failed to recognize is that methodology and measurement are only parts of the process of doing quality research. An equally important aspect is

the use of a theoretical framework, an aspect which is all too easily neglected by an over-emphasis on the techniques of data collection. It is the theoretical framework which guides the ideal research process through the stages of problem specification, hypotheses development, data collection, and interpretation. While the measurement aspect of research has been carried over into evaluation, the guiding role of the theoretical framework has not. Consequently, evaluators often find themselves without adequate guidelines to assist them in formulating criteria and in interpreting the evidence collected by criteria measurements.

In a sense hypotheses are to research what criteria are to evaluation: the controlling dimensions. However, hypotheses emerge out of a theoretical framework in a way that most criteria do not. The results of hypotheses testing are similarly interpreted within that framework. The need for a theoretical framework in the sense that there should be a thorough conceptualization of the evaluation problem is seldom explained in literature presenting the evaluation process. Whether existing theories can be translated into frameworks appropriate for use in evaluation settings, or whether distinctive frameworks and guidelines for their use need to be developed is an interesting area for exploration.

4. <u>Too great an emphasis on information, too little attention to interpreting that information.</u>

With pressure from administration for descriptions of results and with the literature of evaluation stressing the processes of gathering evaluation without emphasizing the role of the conceptual framework in interpreting that evidence, or giving many guides to how the information is to be used, it is no wonder that the evaluator has been seduced into considering evaluation as primarily an information producing activity.

It is disconcerting to see this emphasis even in some of the new literature.[45] Although these authors do indicate that this information may be more useful if it reaches the decision maker in terms of summary statements (probably judgments) the focus is still on the information, not on the use of information in judgment. Until we recognize that information is not the end point in evaluation but merely an input into an end which is the reaching of judgments about programs, we will not feel a need for criteria.

5. The desire to avoid decision making.

A concept of evaluation which puts most of the emphasis on evidence and its processing can be implemented relatively easily by the average person. It requires him to make very few decisions other than a rather general one about possible kinds of evidence to be collected. He accepts givens without challenge. However, a concept which emphasizes criteria spotlights the fact that evaluation involves a sequence of decisions. The programmer and evaluator have to specify the criteria that are to be used and choices are seldom automatic. The person must invest effort to explore alternatives and come to a decision. This is a tiring and sometimes painful effort for many of us. The problem here is very similar to that of the frustration many of us find in writing objectives. The problem is not in putting something on paper but in choosing what it is that we want to put on paper. It commits us to something. It is much easier just to act without dealing with all of the implications that are involved in a deliberate selection and commitment to a visible goal (or in the case of evaluation to a set of standards). Because criteria involve decisions and because decisions are difficult, we tend to busy ourselves with those aspects of evaluation which do not include any difficult decisions.

6. The wish to escape challenge by avoiding subjectivity.

Some people may not use criteria because they recognize them for what they are--subjective decisions. They realize that anything that is thought to have subjective connotations is more open to challenge than is something which appears to be cloaked in objectivity. The mystique of research technology (complicated instruments and sophisticated statistical analysis) makes the whole evidence collection process look objective. It establishes such a smoke screen that the average person lacks the insight to recognize and the courage to challenge the many subjective elements involved. It is more difficult, but still possible, to cloak criteria in such a protective smoke screen. Some people are afraid of challenge. Rather than accepting the challenge of criteria as a means of increasing professional competency and as an important test of validity of assumptions, the coward dodges the value issues involved and retreats to presenting masses of evidence.

Although we feel that the concept of scientific objectivity is very important to evaluation, (in criteria, in evidence and in judgment) we feel that the pendulum swing, initiated to improve the quality of evidence, has gone too far. It has gone so far that "subjectivity" is beyond the pale and subjective realms banned from scientific discussion. This has kept many of the thinking people in our field from helping others develop competency in functioning in subjective situations as objectively as possible. The subjective aspects of program evaluation (and there are many) should be put on the table and dealt with. The process of developing and using criteria is one means of dealing with issues related to the subjective. Kaplan has said that "Freedom from bias means having an open mind, not an empty one."[46]

7. We fail to trust what we know about programming.

Principles are among the soundest sources of criteria. How many principles of adult education do you feel absolutely sure are principles? It would appear that there are relatively few that we are willing to trust sufficiently to use as criteria for judging programs. This may be because we do not have enough proof of the things we believe to be so. We do not know enough about adult education as a total field of study. However, it is ironic that we will operate on a set of beliefs which guide us in carrying out our programs and then be unwilling to state that same set of beliefs as criteria against which to judge the program. Rather than subjecting our guiding beliefs to the scrutiny and challenge that is sure to occur if they are stated as criteria (with the aura of absoluteness that sometimes surrounds criteria) we are more apt to back off and say that there aren't any criteria that can be used. Yet may it not be dangerous to act upon folklore or theory which is not firm enough to serve as criteria?

These are seven of the sources that contribute to the criteria problem in program evaluation.[47] There are others that could be mentioned. Some people feel that criteria are restricting. Some have had poor experiences with using criteria (or more likely, experiences with using poor criteria or with using criteria poorly). As a result, the concept of criteria in program evaluation has been relegated to a remote and dusty corner. We believe, however, that one of the most promising fronts for improving program evaluation and ultimately improving our knowledge of valuable programs for adults is through increased understanding and use of criteria.

Suggested Solutions

Clues to solutions have been inherent in the earlier discussions of

the premises and of the problem. At this point we would like to summarize by discussing the three major avenues through which resolution of the problem can proceed and program evaluation be strengthened by the use of criteria.

1. Theory Development

We would suggest theory development and testing as the first avenue. The development of theory on several fronts can contribute directly and indirectly to improved understanding and use of criteria. First and foremost is the further development of theories of evaluation, particularly theories which include a focus on the role of criteria in evaluation. Specifically, as part of, or as a beginning toward such theoretical frameworks, there is a need for the development of typologies or models which do such things as: help the evaluator understand the phenomena about which he is attempting to develop criteria and help him understand the nature of criteria and how they are developed. Secondly, the development and testing of theories in other areas, such as learning, instruction, and other aspects of the programming process can contribute by crystallizing the principles which should serve as the basis for constant criteria.

2. Research

Research constitutes the second major avenue of attack. In addition to, or as part of research which is carried out specifically to test and further develop the theories mentioned above, the following are suggested as priority research areas related to criteria:

a. There needs to be considerable research on the process of evaluation itself. There have been so few theoretical frameworks to date that much of the evaluative research being done is focusing on the evaluation of the program rather than researching the process of evaluating the program.

b. An important area for research is that which explores the relationships between criteria from various program characteristics. Many of our evaluation procedures which involve criteria other than effectiveness criteria are based on untested (and often unstated) "if-then" propositions. For example, the use of criteria related to program quality may be based on the assumption that "If the quality of learning experiences is high, then the effectiveness of the program will be high." Such propositions need to be tested.

c. Research on the development of criteria, including: the development of constant criteria for use in more than one situation; methods of controling criteria contamination; and the discovery of criteria statement which are most critical and crucial to various kinds of decisions.

d. Research on the process of using criteria to determine such things as the effect of the form of criteria statements; and the degree of specificity which is most efficient and effective.

e. Research on the means of improving the quality of subjective judgment on which criteria selection is ultimately based.

f. Finally there is a need for establishing a pool of results of evaluative research and evaluation studies so that criteria and norms for interpretation can be shared.

3. Strengthening Evaluation Roles.

Leaving theory and research and moving directly into practice, it is our contention that substantial progress can be made on the criteria problem by clarifying and strengthening the roles which administrators, and programmers play in evaluation and by adding a new role--that of evaluation specialists.[48] Adult education agencies whose budgets run

into the millions of dollars can't afford not to hire evaluation specialists (either on their own staffs or on a consulting basis).[49] We are past the time where a positive philosophy and superficial knowledge of procedures on the part of programmers and administrators supplemented by an occasional contract to a social science researcher is sufficient. We need individuals who (1) know adult education well enough to understand the kinds of decisions that need to be made, (2) know the concepts of the field well enough to understand where the crucial points related to those decisions may fall, (3) know the theoretical frameworks and methodology related to a variety of kinds of program intent (e.g. cognitive gain, affective change, environmental change), and (4) can provide thought leadership to the total process of program evaluation.

Given the scarcity of evaluation competencies, it is proposed that the evaluation specialist needs to assume greater responsibility for roles which many have felt properly belong to the programmers, and less responsibility for activities which can be adequately carried out by technical personnel. Specifically we would suggest that the evaluation specialist must take the lead in such areas as:

a. Identifying and involving relevant decision-makers who have a role to play in selecting criteria.

b. Obtaining and assisting with an adequate definition of conceptual criteria.

c. Assisting the programmer to think through such decisions as the types of criteria to be used and the determination of their relative importance.

d. Providing greater input into the criteria development and interpretation processes, based on his familiarity with the literature and other evaluations.

e. Helping administrators use evaluation in budget and other program decisions.

It should be noted that we are not suggesting the evaluation specialist actually make the decisions; rather his role is to assist programmers and administrators to make those decisions and to provide sufficient input to enable them to do so. The point is that competent programmers and administrators cannot also be expected to possess all the competencies needed for evaluation. Systematic evaluation is much too complex a field.

In addition, it is suggested that freeing more of the time of evaluation specialists to work on the above areas can best be accomplished by transferring much of the work of evidence collection and analysis to technical specialists.

There are very few well prepared evaluation specialists in the total field of education and almost none who have adequate background in adult education. In order for the new role of evaluation specialist to emerge as it must in this programming generation, not only must there be an agency commitment, but also there must be a gearing up of graduate programs to prepare such a person. Up until now there have been few professors of adult education specialized in program evaluation. Those who do have a commitment to it have divided their attention between it and other phases of the curriculum process and as a result have done very little research on the evaluation process or little to advance theory. Evaluation needs the undivided attention of some adult education professors working in cooperation with a group of on-the-job specialists who are concentrating their intellectual energies on improving program evaluation. Only then will the break throughs in theory and research be made.

Team Work

A combination of the three avenues--theory, research and specialization--is needed in regard to criteria problems as with other aspects of evaluation. When the resources in these three areas are adequate and the evaluation professor and specialist are able to help the programmer and administrator draw freely upon the principles and criteria built up through the development and testing of evaluation theory then the problems we have outlines will be minimized or disappear. With greater confidence in what is known in the field the programmer and administrator will be able to deal with decisions related to criteria more effectively and be better able to cope with the subjectivity involved.

If the kind of energy involved in developing this paper is any example of the kind of resources that it will take to improve criteria and their use in operationalizing our belief in program evaluation, then it is very clear that progress will require more resources than have presently be allotted. Some will ask whether or not an agency or the field of adult education can afford to allocate these resources. We counter with the question, "How long can it afford not to?", and close with the challenge of this remark of Lewin's:

> In a field that lacks objective standards of achievement, no learning can take place. If we cannot judge whether an action has led forward or backward, if we have no criteria for evaluating the relation between effort and achievement, there is nothing to prevent us from coming to the wrong conclusions and encouraging the wrong work habits.[50]

FOOTNOTES

1. Although Miller sets the stage for a discussion of program criteria by using a quote from Lewin to open his chapter, he reverts to the typical objective and evidence approach and ignores criteria. Miller, Teaching and Learning in Adult Education, The Macmillan Company, 1964, Chapter 10.

   Verner and Booth mention that the use of standards is one of the four principal ways of administrative evaluation of programs, but say that no such standard exists for adult education. They too focus primarily on objectives and of evidence of behavior. Verner and Booth, Adult Education, The Center for Applied Research in Education, Inc., 1964, Chapter 6.

   Thiede draws heavily on Furst (Furst, Constructing Evaluation Instruments, Longmans, Green and Company, 1958) but does not pick-up Furst's comments about criteria of performance. Thiede, "Evaluation and Adult Education," in Jensen, Liveright and Hallenbeck (eds.), Adult Education, Adult Education Association of the U. S. A., 1964, Chapter 15.

   Nor is much attention to criteria per se given in the major Cooperative Extension reference on Evaluation:

   Byrn (ed.) Evaluation in Extension, H. M. Ives and Sons, 1959.

   Sanders, (ed.), The Cooperative Extension Service, Prentice-Hall, Inc., 1966.

   It is to be noted that most of these authors use a similar Tyler-based framework and say about the same thing. They remind one of the eighth grade science text which gives an overview of the world of science for an unsophisticated reader. In science, however, the eighth grade version is backed up by texts at other levels. Adult education authors have not moved beyond writing for the unsophisticated practioner. There are not enough back-up texts of various aspects of programming at a level sufficient to challenge and guide intellectual exploration.

2. Among the recent authors who at least mention criteria are:

   Scriven, "The Methodology of Evaluation," in Tyler, Gagne and Scriven, Perspectives of Curriculum Evaluation, Rand McNally and Company, 1967.

   Stake, "The Countenance of Educational Evaluation, Teachers College Record, Vol. 68, No. 7.

   Stufflebeam, "Toward a Science of Educational Evaluation," Educational Technology, July 30, 1968.

3. Harris, Wilbur, "The Nature and Function of Educational Evaluation," Peabody Journal of Education, September, 1968.

4. This definition is an adaptation of one presented by Steele in Developing a Concept of Program Evaluation, National University Extension Center, University of Wisconsin, February, 1970.

   It is based on a combination of other definitions including:

   A criterion is a standard or rule used to provide a frame of reference for judging or testing something. It is a base, often of a rather arbitrary nature and ultimately involving value judgments, against which comparisons may be made. Ryans, "Notes on the Criterion Problem in Research, with Special Reference to Teacher Effectiveness," The Journal of Genetic Psychology, XCI (September, 1957).

   A criterion is "a comparison object, or a rule, standard or test for making a judgment . . . a behavior goal by which progress is judged . . . the variable, comparison with which constitutes a measure of validity." English and English, A Comprehensive Dictionary of Psychological and Psycholanalitical Terms. Longmans, Green, 1958, as cited in Astin, "Criterion-Centered Research," Educational and Psychological Measurement, Vol. XXIV, No. 4, 1964.

   A criterion is "a behavior or condition which is or can be described in terms of an ideal . . . a goal . . . behavior which is considered desirable and toward which one works." Jensen, Coles and Nestor, "The Criterion Problem in Guidance Research," Journal of Counseling Psychology, II (1955), as cited by Astin, op. cit.

   A criterion is a rule by which values are assigned to alternatives, and optimally such a rule includes the specification of variables for measurement and standards for use in judging that which is measured. Stufflebeam, op. cit.

5. Recently Scriven has striven to emphasize the difference between the goal of evaluation (that of answering certain types of questions) and the various roles of evaluation. Scriven, op. cit., p. 40.

   Two centers, the Ohio State University Evaluation Center and the UCLA Center for the Study of Evaluation are putting a high degree of emphasis on evaluation's role in decision making.

   Although Extension and Adult Education have given lip service to the concept for some time, most of the ideas related to the process of evaluation focus on sumnative evaluation which has not often been used or usable in important programming decisions.

6. Stufflebeam, op. cit., p. 6, see also: Alkin, "Evaluation Theory Development," UCLA Evaluation Comment, October, 1969, Vol. 2, No. 1.

7. Alkin, op. cit., p. 3.

8. Stufflebeam lists the extent to which the evaluation is trusted as one of the criteria for evaluating evaluation. op. cit., p. 6.

9. See the September, 1969 issue of The Annals of the American Academy of Political and Social Science. In particular: Ferman "Some Perspectives on Evaluating Social Welfare Programs;" Weiss and Rein "The Evaluation of Broad-Aim Programs: A Cautionary Case and A Moral;" Williams and Evans, "The Politics of Evaluation: The Case of Head Start." Although written from the standpoint of situations where independent units do the evaluation, the problems described are also relevant in other situations.

10. Stake, op. cit., p. 535.

11. Stake refers to description and judgment as the two basic acts of evaluation and states: "To be fully understood, the educational program must be fully described and fully judged." Stake, op. cit.

    Gruba feels that "most evaluators agree that mere collection of data does not constitute evaluation." Gruba, "The Failure of Educational Evaluation," Educational Technology, May, 1969.

12. Steele, op. cit.

    Most current theory of program or curricula evaluation seem to become so emmeshed with one or more dimensions of program that it gives inadequate attention to the theoretical components of evaluation. Stake is an exception.

13. We may be differing somewhat from Stake on this. He seems to give the impression that the two programs are compared directly without recourse to criteria statements. Stake, op. cit., p. 536, 538.

14. We will not take space here to explain the categories in any detail. See "Toward A Typology of Program Criteria," working paper by Steele and Moss, Department of Agricultural and Extension Education, University of Wisconsin, 1970, for more detail.

15. Stufflebeam includes "context" and Stake "antecedent" in their evaluation frameworks. As far as we can understand the concepts they deal primarily or at least in part with factors prior to the program. Stufflebeam, op. cit.; Stake, op. cit.

16. Scriven's terms; op. cit., p. 43.

17. Similar to Scriven's term "explanatory" evaluation. He seems to have doubts about the appropriateness of this kind of evaluation. We feel that it is an essential type of evaluation. It differs from evaluative research in that evaluation searches for answers specific to a given program; research searches for conclusions that can be generalized to more than the one programming situation.

18. See such sources as:

    Furst, op. cit., p. 57-79.

    Mager, Preparing Objectives for Programmed Instruction, Fearon Publishers, 1962.

Burns, "The Theory of Expressing Objectives," Educational Technology, October 30, 1967.

19. Scriven, op. cit., p. 62-66; Sawin, Evaluation and the Work of the Teacher, Wadsworth Publishing Company, Inc., 1969.

20. Scriven, op. cit., p. 77-80.

21. Herzog, Elizabeth. Some Guide Lines for Evaluative Research, U. S. Department of Health, Education and Welfare, Children's Bureau, Washington, D. C., 1959, p. 17.

22. See Suchman, op. cit., and Byrn, op. cit.

23. Suchman, op. cit.

24. Ryans, op. cit., p. 39.

25. Astin, op. cit., p. 810.

26. Stake, op. cit., p. 538, and Ryans, op. cit., p. 39.

27. See for example, Ryans, op. cit., p. 34-35, and Stake, op. cit., p. 536.

28. Astin, op. cit., p. 814.

29. Ryans, op. cit.

30. Ryans, op. cit., p. 35. The shortcomings of this approach are also pointed out by Brogden and Taylor in their criticism of the use of available criterion measures as the basis of criteria development. Brogden and Taylor, op. cit., p. 162.

31. Ryans, op. cit., p. 36.

32. Ibid.

This approach is similar to Brogden and Taylor's first step in criterion construction which they describe as consisting of: "Careful analysis of the total situation in which the criterion behavior occurs for the purpose of isolating all sub-criterion variables and obtaining preliminary estimates of their relative importance." Brogden and Taylor, op. cit., p. 162.

33. Tyler, Ralph W. Basic Principles of Curriculum and Instruction. The University of Chicago Press, Chicago, Illinois, 1950.

34. Ryans, op. cit., p. 38.

35. Astin notes that "Perhaps the most common misconception about criterion measures is the notion that they can be 'validated' . . . the only method for 'validating' a criterion measure is a logical

analysis of its relevance to the conceptual criterion." Astin, op. cit., p. 811. See also Brogden and Taylor, op. cit., p. 160.

36. Ryans, op. cit., p. 42-44.

37. Suchman, op. cit., p. 120.

38. Brogden and Taylor, op. cit.

39. Scriven, op. cit., p. 57.

40. For example, Brogden and Taylor refer to the need to obtain "preliminary estimates" of the relative importance of criteria, while Ryans refers to the need to take into account the relative importance of each component of a dimension and of each dimension contributing to the overall criterion. Brogden and Taylor, op. cit., p. 162; Ryans, op. cit., p. 39.

41. Brogden and Taylor, op. cit., p. 183-184.

42. Scriven states that "we may often wish to alter the weighting of a variable when it drops below a certain level." Scriven, op. cit., p. 73.

43. Tyler, Basic Principles of Curriculum and Instruction, University of Chicago Press, 1950.

44. Gruba, op. cit.

45. For example, Alkins, op. cit., and Stufflebeam, op. cit.

46. Kaplan, The Conduct of Inquiry, Chandler Publishing Company, 1964, p. 357.

47. These seven sources have been presented within the broad concept of evaluation which we have developed in this paper. Most are equally applicable to the more narrow traditional approach which emphasizes attainment of objectives: (1) we have not fully understood what it means to evaluate the extent to which objectives have been attained; (2) we have concentrated too much on evidence of behavior and not enough on what the evidence means; (3) we have avoided decision-making at the stage of setting objectives and have not been willing or able to commit ourselves to a precise definition of what is expected; (4) we emphasize the compilation of "objective" data in order to avoid those aspects which may be subjective; (5) we fail to use a conceptual framework; (6) we use results of the evaluation more often for reports than for making decisions on programming methods; and finally (7) we simply do not know enough (or trust sufficiently what we do know) about learning and behavioral change to set specific criteria for the performance of our program participants.

48. For one taxonomy of evaluator tasks and roles see Owens, Thomas R. "Suggested Tasks and Roles of Evaluation Specialist in Education," Educational Technology, November 30, 1968, p. 4-10.

49. We will not debate the wisdom of the evaluation specialist being employed within an agency or being contracted for from an independent source. We feel there is a need for both if the evaluation is to be trusted by relevant decision makers. Certainly there is a need within an agency for an evaluation consultant to work closely with programmers in the continual complex process of program evaluation and to help administration understand and make use of the judgments developed in cooperation with the programmers. When large amounts of money or considerable prestige is at stake the agency may find it best to contract certain types of evaluation out to a consultant. In this case, we feel that it should be a "full time evaluator" conversant with the field of adult education rather than to a researcher who occasionally takes on an evaluation project in order to achieve certain research or financial ends.

50. Lewin, Kurt. "Group Decision and Social Change," as cited by Miller, op. cit.

ERIC Clearinghouse
FEB 25 1970
on Adult Education